GEOMETRIC ERGODICITY OF TWO–DIMENSIONAL HAMILTONIAN SYSTEMS WITH A LENNARD–JONES–LIKE REPULSIVE POTENTIAL

BEN COOKE, JONATHAN C. MATTINGLY, SCOTT A. MCKINLEY,
AND SCOTT C. SCHMIDLER

1. Introduction

Molecular dynamics simulation is among the most important and widely used tools in the study of molecular systems, providing fundamental insights into molecular mechanisms at a level of detail unattainable by experimental methods [AT87, Lea96, FS96, Sch02]. Usage of molecular dynamics spans a diverse array of fields, from physics and chemistry, to molecular and cellular biology, to engineering and materials science. Due to their size and complexity, simulations of large systems such as biological macromolecules (DNA, RNA, proteins, carbohydrates, and lipids) are typically performed under a classical mechanics representation. A critical requirement of such simulations is ergodicity, or convergence in the limit to the equilibrium (typically canonical) Boltzmann measure $\mu(d\mathbf{q}, d\mathbf{p}) = Z(\beta)^{-1}\mathrm{e}^{-\beta H(\mathbf{q},\mathbf{p})}d\mathbf{q}d\mathbf{p}$. Although ergodicity is commonly assumed, recently [CS08] showed that all commonly used deterministic dynamics methods for simulating the canonical (constant-temperature) ensemble fail to be ergodic. They also showed that introduction of a stochastic hybrid Monte Carlo (HMC) correction guarantees ergodicity; however, HMC scales poorly with system dimension and is rarely used for macromolecules. [CS08] also show empirically that more commonly used stochastic Langevin dynamics [Pas94] appear to exhibit ergodic behavior, but were unable to provide rigorous proof.

The key difficulty in applying existing arguments [MSH02] is the appearance of singularities in the potential $U(\mathbf{q})$. Most modern molecular mechanics forcefields [PCC+95, BBO+83, JTR88] take the form

$$U(\mathbf{q}) = \sum_{\text{bonds}} K_1(r - r^*)^2 + \sum_{\text{angles}} K_2(\theta - \theta^*)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\epsilon\, r_{ij}}\right]$$

Here the first three terms involve bond length, angle, and torsional energies; being bounded, these are easily handled. The difficulty arises from the non-covalent electrostatic and Van der Waals forces, the latter modeled by a Lennard-Jones potential, which give rise to singularities as two atoms in the system approach each other at close range.

Date: January 20, 2013.

In this paper we establish the ergodicity of Langevin dynamics for simple two-particle system involving a Lennard-Jones type potential. To the best of our knowledge, this is the first such result for a system operating under this type of potential. Moreover we show that the dynamics are *geometrically* ergodic (have a spectral gap) and converge at a geometric rate. Geometric ergodicity is sufficient to imply existence of a central limit theorem for ergodic averages of functions h with $\mathbf{E}_\mu(|h|^{2+\delta}) < \infty$ for some $\delta > 0$ [IL71], and also implies the existence of an exact sampling scheme [Ken04], although the latter need not be practical. Loosely, proving an ergodic result has two central ingredients. One provides continuity of the transition densities in total variation norm which ensures that transitions from nearby points behave similarly enough probabilistically which provides the basic mechanism of the probabilistic mixing/coupling. This is often expressed in a minorization condition (see Section 6). The other ingredient gives control of excursions towards infinity which ensures the existence of a stationary measure and guarantees that sufficient probabilistic mixing for an exponential convergence rate. The difficulty in a problem is typically one or the other.

In Section 2, we will see that in the current setting, basic existence of a stationary measure is trivial since the standard Gibbs measure built from the energy is invariant. Uniqueness of the stationary distribution follows from now standard results on hypoelliptic diffusion. However the control necessary to give a convergence rate has previously been elusive. Our approach follows the established method of demonstrating the existence of a Lyapunov function and associated small set; however, construction of the Lyapunov function in the presence of a singular potential is non-trivial and our approach constitutes one of the major innovations of this paper. In many ways it builds on ideas in [HM09] and more obliquely is related to the ideas in [RBT02]. In both cases, time averaging of the instantaneous energy dissipation rate is used to build a Lyapunov function. We use similar ideas here. In a nutshell, as in [HM09] the technique consists of casting the behavior of the system as the energy heads to infinity as a problem with order one energy containing a small parameter equal to one over the original systems energy. Then classical stochastic averaging techniques are used to build a Lyapunov function. Though the solution is related to [HM09], the presentation of difficulties is quite different. We conclude by briefly discussing the challenges of extending our results to larger systems and the case of harmonically growing potential which are not covered by our theorems.

2. A Model Problem

We will use (Q, P) to denote the position and momentum of a particle in a deterministic system, and (q, p) for the corresponding stochastic system.

Consider the two-particle Hamiltonian system $(\mathbf{Q}, \mathbf{P}) = ((Q_1, Q_2), (P_1, P_2))$ with Hamiltonian

$$H_0(\mathbf{Q}, \mathbf{P}) = \frac{P_1^2}{2} + \frac{P_2^2}{2} + U(Q_1 - Q_2)$$

and interaction potential

$$U(R) = \sum_{k=1}^{K} a_k |R|^{\alpha_k} + a_0 > 0 \tag{1}$$

where $a_k \in \mathbb{R}$ with $a_1, a_K > 0$, and $\alpha_1 > \cdots > \alpha_K$. We assume that $\alpha_1 > 2$ and $\alpha_K < 0$ (otherwise no singularity exists).

The dynamics of this system are given by

$$\dot{Q}_i = \frac{\partial H_0}{\partial P_i} \qquad \dot{P}_i = -\frac{\partial H_0}{\partial Q_i} \qquad \text{for } i = 1, 2.$$

If we force the system with a noise whose magnitude is scaled to balance dissipation so as to place the system at temperature T then we arrive at the system of coupled SDEs

$$\begin{aligned} dq_i &= p_i\, dt \qquad \text{for } i = 1, 2 \\ dp_1 &= -U'(q_1 - q_2)\, dt - \gamma p_1\, dt + \sigma dW_1(t) \\ dp_2 &= \ U'(q_1 - q_2)\, dt - \gamma p_2\, dt + \sigma dW_2(t) \end{aligned} \tag{2}$$

where $\sigma^2 = 2\gamma T$. Define

$$\mathbb{S} \stackrel{\text{def}}{=} \Big\{ (p_1, q_1, p_2, q_2) : q_1 \neq q_2 \Big\}.$$

We will prove in Corollary 5.4 that, if the initial conditions are in $\mathbb{S}$, then with probability one there exists a unique strong solution to (2) which is global in time and stays in $\mathbb{S}$.

We define the Markov semigroup by $\mathcal{P}_t \phi(\mathbf{p}, \mathbf{q}) \stackrel{\text{def}}{=} \mathbf{E}_{(\mathbf{p},\mathbf{q})} \phi(\mathbf{p}_t, \mathbf{q}_t)$ where $\mathbf{E}_{(\mathbf{p},\mathbf{q})}$ is the expected value starting from $(\mathbf{p}, \mathbf{q})$. This semigroup has a generator $\mathcal{L}_0$ given by

$$\mathcal{L}_0 \stackrel{\text{def}}{=} \sum_{i=1,2} \frac{\partial H_0}{\partial p_i} \frac{\partial}{\partial q_i} - \frac{\partial H_0}{\partial q_i} \frac{\partial}{\partial p_i} - \gamma p_i \frac{\partial}{\partial p_i} + \gamma T \frac{\partial^2}{\partial p_i^2}\,.$$

Additionally $\mathcal{P}_t$ induces a dual action on σ-finite measures μ by acting on the left: $\mu \mathcal{P}_t$. A measure μ_0 is a stationary measure of $\mathcal{P}_t$ if $\mu_0 \mathcal{P}_t = \mu_0$. In our setting, this is equivalent to asking that $\mathcal{L}_0^* \rho_0 = 0$ where $\mu_0(d\mathbf{p}, d\mathbf{q}) = \rho_0(\mathbf{p}, \mathbf{q}) d\mathbf{p}\, d\mathbf{q}$.

It is a simple calculation to see that if

$$\rho_0(\mathbf{p}, \mathbf{q}) \stackrel{\text{def}}{=} C e^{-H_0(\mathbf{p},\mathbf{q})/T}$$

for some C, then $\mathcal{L}_0^* \rho_0(\mathbf{p}, \mathbf{q}) = 0$. Hence with this choice of ρ_0, μ_0 as defined above is a stationary measure. However this measure is not normalizable to make a probability measure since it is only σ-finite. This stems from the fact that the Hamiltonian is translationally invariant in $\mathbf{q}$. To rectify his problem we will move to "center of mass" coordinates.

2.1. **Reduction to Center of Mass Coordinates.** Let $\tilde{q} = \frac{1}{2}(q_1 - q_2)$, $\tilde{p} = \frac{1}{2}(p_1 - p_2)$, $\bar{q} = \frac{1}{2}(q_1 + q_2)$, $\bar{p} = \frac{1}{2}(p_1 + p_2)$, $W = \frac{1}{2}(W_1 - W_2)$ and $B = \frac{1}{2}(W_1 + W_2)$. Then

$$\begin{aligned} d\bar{q}_t &= \bar{p}_t\, dt \\ d\bar{p}_t &= -\gamma \bar{p}_t\, dt + \sigma dB_t \\ d\tilde{q}_t &= \tilde{p}_t\, dt \\ d\tilde{p}_t &= -U'(2\tilde{q}_t)\, dt - \gamma \tilde{p}_t\, dt + \sigma dW_t \end{aligned} \tag{3}$$

In these new coordinates, the system is described by variables $(\bar{q}, \bar{p})$ tracking the position and momentum of the center of mass, and variables $(\tilde{q}, \tilde{q})$ tracking the relative position and momentum of the particles within the center of mass frame.

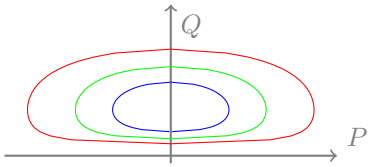

FIGURE 1. Level sets of $H(Q,P) = \eta$ for η equals 1 (in blue), 2 (in green), and 4 (in red) where $H(Q,P) = \frac{1}{2}P^2 + Q^4 + \frac{1}{10}Q^{-2}$.

This change of coordinates simplifies our problem to two uncoupled Hamiltonian sub-problems. The center of mass $(\bar{q}, \bar{p})$, has Hamiltonian

$$\bar{H}(\bar{q}, \bar{p}) \stackrel{\text{def}}{=} \frac{\bar{p}^2}{2}$$

which is the Hamiltonian of a free 1D particle, with corresponding invariant measure given by a Gaussian (for momentum $\bar{p}$) times 1D Lebesgue measure (for position $\bar{q}$). Note that $\bar{p}$ follows an Ornstein-Uhlenbeck process and hence converges exponentially quickly to its (Gaussian) stationary measure. The position $\bar{q}$ will diffuse through space like 1D Brownian motion and hence converges to Lebesgue measure.

The remaining two variables $(\tilde{q}, \tilde{p})$ are also a Hamiltonian system with Hamiltonian

$$H(\tilde{q}, \tilde{p}) \stackrel{\text{def}}{=} \frac{\tilde{p}^2}{2} + U(2\tilde{q})\,. \tag{4}$$

which is a single particle interacting with a potential U that is attractive towards the origin at large distances, and repulsive at short distance. So $(\tilde{q}, \tilde{p})$ will have an invariant *probability* measure. However convergence of this system is more subtle; it possesses two difficulties stemming from the structure of the potential. First, since $U(R)$ is singular at points, a strictly positive density does not exist everywhere in space. Second, there is no immediate candidate for a Lyapunov function. Overcoming this second obstacle will prove more difficult and will occupy the bulk of this paper.

3. Reduced system : main results

We now turn to the study of the two–dimensional Hamiltonian system described by (4). In this section we state the principal results on this reduced system.

Consider the two-dimensional deterministic Hamiltonian system with Hamiltonian

$$H(Q,P) \stackrel{\text{def}}{=} \frac{P^2}{2} + U(Q)$$

and hence dynamics

$$\dot{Q}_t = \frac{\partial H}{\partial P}(Q_t, P_t) = P_t \qquad \text{and} \qquad \dot{P}_t = -\frac{\partial H}{\partial Q}(Q_t, P_t) = -U'(Q_t)\,.$$

This system has only closed orbits, which lie completely in the upper half plane denoted by $\mathbb{H} = \{(Q,P) \in \mathbb{R}^2 : Q > 0\}$ provided the initial points lie in $\mathbb{H}$. To see this observe that when $|(Q,P)| \to \infty$, $H(Q,P)$ is well approximated by $\frac{1}{2}P^2 + a_1 Q^{\alpha_1} + a_K Q^{\alpha_K}$ which clearly has level sets that are closed, homotopically a circle, and lie completely in the upper half plane. (See Figure 1).

Addition of balanced noise and dissipation yields the associated stochastic system of interest. Namely, for positive temperature T, friction γ and noise standard deviation $\sigma = \sqrt{2\gamma T}$, we have

$$\begin{aligned} dq_t &= p_t\,dt \\ dp_t &= -U'(q_t)\,dt - \gamma p_t\,dt + \sigma\,dW_t \end{aligned} \tag{5}$$

This Markov process has generator

$$\mathcal{L} = \frac{\partial H}{\partial p}\frac{\partial}{\partial q} - \frac{\partial H}{\partial q}\frac{\partial}{\partial p} - \gamma p\frac{\partial}{\partial p} + \gamma T\frac{\partial^2}{\partial p^2}$$

and as in the previous section a straightforward calculation shows that $\mu_*(dp\times dq) = \rho_*(q,p)dpdq$ is a stationary measure with

$$\rho_*(q,p) = Ce^{-H(q,p)/T},$$

since $\mathcal{L}^*\rho_* = 0$. Unlike the stationary measure of the unreduced system, this measure can be normalized and made into a probability measure (since H is no longer translationally invariant) for an appropriate choice of C.

In fact ρ_* is the unique stationary measure of the system. To see this first observe that (5) is hypoelliptic and hence any weak solution to $\mathcal{L}^*\mu = 0$ must locally have a smooth density with respect to Lebesgue measure. Since μ_* has an everywhere positive density with respect to Lebesgue measure it must therefore be the only stationary measure, since any stationary measure can be decomposed into its ergodic components all of which must have disjoint support. Uniqueness of the stationary measure is also a by-product of the exponential convergence given in Theorem 3.1 which is our principal interest here.

To state this convergence result we need a distance between probability measures appropriate for our setting. To this end, for any $\beta \geq 0$ we define for $\phi : \mathbb{H} \to \mathbb{R}$ the weighted supremum-norm

$$\|\phi\|_\beta \stackrel{\text{def}}{=} \sup_{(q,p)\in\mathbb{H}} |\phi(q,p)|e^{-\beta H(q,p)}$$

and the weighted total-variation norm on signed measure ν (with $\nu(\mathbb{H}) = 0$) by

$$\|\nu\|_\beta \stackrel{\text{def}}{=} \sup_{\phi:\|\phi\|_\beta\leq 1} \int_{\mathbb{H}} \phi\,d\nu\,.$$

When $\beta = 0$ this is just the standard total-variation norm. We define $\mathcal{M}_\beta(\mathbb{H})$ to be the set of probability measures μ on $\mathbb{H}$ with $\int_{\mathbb{H}} \exp(\beta H)d\mu < \infty$. Then we have the following convergence result

Theorem 3.1. *There exists $\beta_0 > \frac{2}{T}$ so that for any $\beta \in (0,\beta_0)$ there exist positive constants C and δ such that for any two probability measures $\mu_1, \mu_2 \in \mathcal{M}_\beta(\mathbb{H})$*

$$\|\mu_1\mathcal{P}_t - \mu_2\mathcal{P}_t\|_\beta \leq Ce^{-\delta t}\|\mu_1 - \mu_2\|_\beta$$

In particular the system has a unique invariant measure, which necessarily coincides with μ_* defined above, and to which the distribution of (q_t, p_t) converges.

Remark 3.2. *We will see later that β_0 from Theorem 3.1 equals $2\Lambda_*/T$ where Λ_* is the constant defined in* (12) *which depends only on the choice of α_1 in the potential U but always satisfies $\Lambda_* \in (1,2)$.*

Our proof of Theorem 3.1 will follow the now standard approach of establishing the existence of an appropriate "small set" and Lyapunov function [MT93]. Similar to [MSH02], we will use a control argument coupled with hypo-ellipticity to establish the existence of a small set. While this is rather standard, the technique used to prove the existence of a Lyapunov function is less standard and one of the central contributions of this note.

4. The Lyapunov function: Overview

4.1. **Heuristics and motivating discussion.** We wish to control motion out to infinity ($|(q,p)| \to \infty$) as well as in the neighborhood of the singularity ($q \to 0$). A standard route to obtaining such control is to find a Lyapunov function $V : \mathbb{H} \to (0,\infty)$ so that if $V_t = V(q_t, p_t)$ then

$$dV_t \leq -cV_t dt + Cdt + dM_t \tag{6}$$

for some martingale M_t and positive constants c, C and such that $H \leq c_0 V$ for some positive c_0.

The first reasonable choice might be to try the Hamiltonian $H(q,p)$ itself. Setting $H_t = H(q_t, p_t)$ we see that

$$dH_t = -\gamma p_t^2 dt + \frac{\sigma^2}{2} dt + \sigma p_t dW_t \,. \tag{7}$$

However the function $(q,p) \mapsto p^2$ is not bounded from below by $(q,p) \mapsto H(q,p)$ since the two functions are not comparable. This prevents us from obtaining the desired bound. If $U(R)$ only has positive powers of R which are greater or equal to two, this deficiency can be partially overcome by considering $V_t = H_t - \gamma_0 p_t q_t$. Then by picking γ_0 small enough, we can ensure that $\frac{1}{c}H \leq V \leq cH$ as $p^2+q^2 \to \infty$ and that $\mathcal{L}V$ is bounded from above by a constant times $-V + C$ for some $C > 0$. Hence V is comparable to H but satisfies the desired Lyapunov-function inequality (6). See [MSH02] for more on using this trick in this context.

Unfortunately, this simple trick does not seem to work in the presence of a singular repulsive term, as it does not address the explosion of H as q approaches 0 though it does rule out explosion as $|(q,p)| \to \infty$. This is necessary since the potential and hence the transition density behaves poorly near this point and uniform estimates are not easy (if possible) to obtain. Eventually we will find an appropriate function Ψ so that $V = H + \Psi$ works; to do so we need to better understand the dynamics.

With this in mind, we return to (7) and take a closer look at the dynamics. Looking at the right hand side, it is true that p^2 is not comparable to $H(q,p)$ at every given point (q,p) in phase space. Yet if we really believe that the system does not blowup the $-p^2$ must lead to some "dissipation" of energy when the energy is large.

Since we are only interested in preventing the blowup of H, we need only consider the equation when the energy is large. However, when the energy is large we know a lot about the dynamics. To leading order it will follow the deterministic dynamics with stochastic fluctuations of lower order. At high energy, the highest order part of the potential U dominates.

For discussion purposes, in this section, we will assume that

$$U(R) = a_1|R|^{\alpha_1} + a_2|R|^{\alpha_2} \tag{8}$$

for some $a_1, a_2 > 0$ and $\alpha_1 > 2 > 0 > \alpha_2$. Later in Section 5, we will return to the general problem. It will be convenient to introduce the following family of potentials indexed by a parameter $\lambda \geq 1$

$$U(R;\lambda) = a_1|R|^{\alpha_1} + a_2|R|^{\alpha_2}\lambda^{2(\frac{\alpha_2}{\alpha_1}-1)}.$$

Setting $\lambda = 1$ yields the original problem. The advantage to considering this family of potentials is that $U(R;\lambda)$ has the following homogeneity property for $\ell > 0$

$$U(\ell^{\frac{2}{\alpha_1}}R;\lambda) = \ell^2 U(R;\ell\lambda)$$

which will lead to all of the scaling properties mentioned subsequently.

The orbits of the deterministic trajectories are given by the solution set of $H(Q,P;\lambda) = \frac{1}{2}P^2 + U(Q;\lambda) = \eta$ for a given energy level $\eta > 0$. This locus is topologically equivalent to a circle and hence setting

$$\varrho(Q;\eta,\lambda) = \sqrt{2(\eta - U(Q;\lambda))}, \tag{9}$$

the orbit is given by the set $\{(Q,\varrho(Q;\eta,\lambda)),(Q,-\varrho(Q;\eta,\lambda)) : Q \in [Q_-(\eta,\lambda), Q_+(\eta,\lambda)]\}$ where $Q_+(\eta,\lambda)$ and $Q_-(\eta,\lambda)$ are respectively the largest and smallest positive roots of $\eta - U(Q;\lambda) = 0$.

We wish to determine the average of P^2 around the deterministic orbits, as this will give a good idea of the dissipation asymptotically as the energy becomes large. We see that averaging P^2 around this deterministic trajectory gives

$$\langle P^2\rangle(\eta,\lambda) = 2\int_{Q_-(\eta,\lambda)}^{Q_+(\eta,\lambda)} \varrho(Q;\eta,\lambda)dQ\,;$$

and similarly that the period $\tau(\eta,\lambda)$ of this orbit can be expressed as

$$\tau(\eta,\lambda) = 2\int_{Q_-(\eta,\lambda)}^{Q_+(\eta,\lambda)} \frac{1}{\varrho(Q;\eta,\lambda)}dQ$$

To make the idea of "large energy" more precise we consider the rescaling of phase space defined by the mapping $(q,p) \mapsto (\ell P, \ell^{\frac{2}{\alpha_1}}Q)$ for a scale factor $\ell > 0$. Under this map the associated energy will essentially scale by a factor ℓ^2 for large ℓ. However this is not *exactly* correct since the other terms in the potential do not scale in the same fashion. However by changing the value of λ we can relate a scaled Hamiltonian exactly with an unscaled Hamiltonian having $\lambda = \frac{1}{\ell}$; that is, since $H(Q,P;\lambda) = \frac{1}{2}P^2 + U(Q;\lambda)$, we see that $H(\ell P, \ell^{\frac{2}{\alpha_1}}Q;\lambda) = \ell^2 H(Q,P;\ell\lambda)$. In other words, the scaled system behaves exactly like the unscaled system at a higher energy. If we define the average

$$\mathcal{A}(P^2)(\eta,\lambda) \stackrel{\text{def}}{=} \langle P^2\rangle(\eta,\lambda)/\tau(\eta,\lambda) \tag{10}$$

then we also see that $\mathcal{A}(P^2)(h\eta,\lambda) = h\mathcal{A}(P^2)(\eta, h^{\frac{1}{2}}\lambda)$.

Summarizing, the average of P^2 around the deterministic orbit with energy $h\eta$ and $\lambda = 1$ is the same as h times the average of P^2 around the deterministic orbit with energy η and $\lambda = h^{\frac{1}{2}}$ for the simplified potential in this section. We will see that this will hold for sufficiently large energy in the more general setting discussed in Section 5. If we define,

$$\Lambda(\eta) \stackrel{\text{def}}{=} \mathcal{A}(P^2)(1;\eta^{\frac{1}{2}}) \tag{11}$$

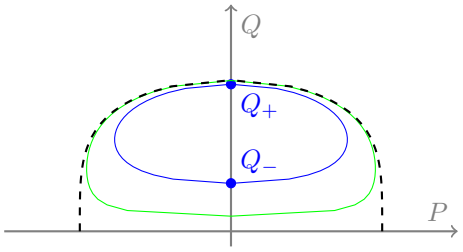

FIGURE 2. Level sets of $H(Q,P;\lambda)=1$ for λ equals 1 (blue) and 2.15 (green) where $H(Q,P;\lambda)=\frac{1}{2}P^2+Q^4+\frac{1}{10}\lambda^{-3}Q^{-2}$ The dashed line is the level set of $\frac{1}{2}P^2+Q^4=1$ with $P\geq 0$ to which the level sets of $H(Q,P;\lambda)=1$ converge as $\lambda\to\infty$.

then $\mathcal{A}(P^2)(\eta,1)=\eta\Lambda(\eta)$. Further observe that as $\lambda\to\infty$, the level sets under potential $U(R;\lambda)$ converge (Figure 2), and $\mathcal{A}(P^2)(1;\lambda)$ converges to a positive constant Λ_*. As we will see later

$$\Lambda_* = \frac{\int_0^{Q_*}\left(1-a_1Q^{\alpha_1}\right)^{\frac{1}{2}}dQ}{\int_0^{Q_*}\left(1-a_1Q^{\alpha_1}\right)^{-\frac{1}{2}}dQ} = \frac{2\alpha_1}{\alpha_1+2} \tag{12}$$

where $Q_*=a_1^{-\frac{1}{\alpha_1}}$. Notice that Λ_* is independent of the value of a_1 and since $\alpha_1>2$, observe that $\Lambda_*\in(1,2)$.

Now since at high energy (i.e. $\eta\gg 1$), $\mathcal{A}(P^2)(\eta,1)=\eta\Lambda(\eta)\approx\eta\Lambda_*$, it is reasonable to approximate (7) by

$$dH_t \approx -\gamma\Lambda_* H_t dt + \frac{\sigma^2}{2}dt + \sigma\sqrt{\frac{\Lambda_* H_t}{2}}dW_t \tag{13}$$

when $H_t\gg 1$. In making this approximation, we are *not* claiming that there is averaging in the traditional asymptotic sense. Namely, that there is a small parameter going to zero which causes the system to speed up and hence the instantaneous effect on the system is increasing in the limit that of the averaged parameter. Rather, at high energy the system acts (after rescaling) increasingly like a system with order one energy and a rescaled parameter λ. The rescaling also leads to a rescaling of time so that an order one time in the rescaled system represents an increasingly short time in the original system. Hence in a short instance of time at high energy, one sees the effect of many rotations of the system, making the averaged quantities just calculated a good approximation.

4.2. **Numerical explorations.** The plots in Figure 3 compare the trajectory of the energy predicted by (13) and the energy trajectory obtained from a numerical simulation of (5) when both were started from the same initial high energy level. The model potential given in (8) was used with $\alpha_1\in\{2,4,6\}$ and $\alpha_2=-12$. Similar comparisons with α_2 equal to -2 and -4 were also made with nearly identical plots confirming essentially no dependence on α_2 as predicted by our asymptotic theory.

Our theory only applies to the two cases $\alpha_1\in\{4,6\}$. In these cases the agreement with the theory, shown with the dashed line, is quite good. One can see a small scale wiggle in the numerical curves. This is the effect of the periodic orbit. As the scaling theory predicts, the effect decreases as the energy increases since the scaling

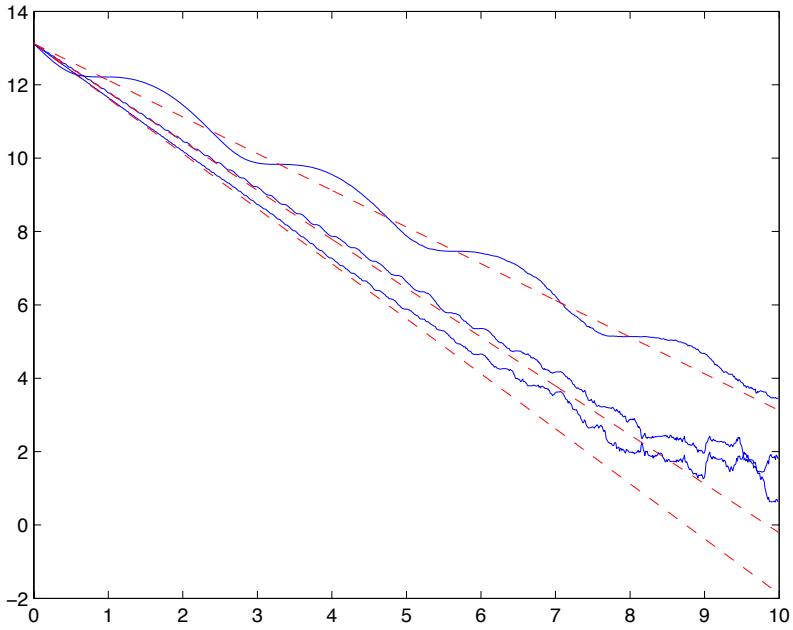

FIGURE 3. The first three plots are semi-log plots of energy versus time for the dynamics using the potential in (8) with α_1 equal to 2 (uper most curve), 4 (middle curve), and 6 (lower most curve).The solid lines are numerical simulations and the dashed lines are the theoretical prediction made by (13).

shows that period and the size of the fluctuations go to zero as the energy increases. When $\alpha_1 = 2$ our theory does not apply. Nonetheless, the trend given by dotted line is followed. However one sees that period and amplitude of the fluctuation is not going to zero which is also constant with the scaling arguments predictions. The possibility of extending our theory to this boundary case is discussed in Section 8.

4.3. **Using the Poisson equation: Overview.** Informed by the preceding discussion, we return to the idea of constructing a Lyapunov function V of the form $V = H + \Psi$, where Ψ is introduced to handle the singularity in H. To see how the average value of $-p^2$ comes into this framework and make rigorous the intuition from (13), we will make use of the "Poisson equation" associated with the deterministic dynamics, which controls the fluctuation of $\mathcal{A}(P^2)$.

We begin by introducing the Liouville operator $\mathcal{H}$ associated with the deterministic dynamics and defined by

$$(\mathcal{H}\phi)(Q,P;\lambda) \stackrel{\text{def}}{=} \frac{\partial H}{\partial P}(Q,P;\lambda)\frac{\partial \phi}{\partial Q}(Q,P) - \frac{\partial H}{\partial Q}(Q,P;\lambda)\frac{\partial \phi}{\partial P}(Q,P) \tag{14}$$

for $\phi : \mathbb{R}^2 \to \mathbb{R}$. Next we define Ψ to be the solution to the Poisson equation

$$(\mathcal{H}\Psi)(Q,P;\lambda) = \gamma\big(P^2 - \mathcal{A}(P^2)(Q,P;\lambda)\big) \tag{15}$$

where $\mathcal{A}(P^2)(\eta,\lambda)$ is the averaging operator defined by (10) discussed in Section 4.1 and we have introduced the slight abuse of notation

$$\mathcal{A}(P^2)(Q,P;\lambda) \stackrel{\text{def}}{=} \mathcal{A}(P^2)(H(Q,P;\lambda);\lambda)\,. \tag{16}$$

Since $P \mapsto P^2$ is a smooth function, we will see that (15) has a smooth solution.

As mentioned, we will choose the Lyapunov function $V(q,p) = H(q,p) + \Psi(q,p)$. To understand why this is a good choice, first observe that the generator $\mathcal{L}$ of (5) can be written as

$$\mathcal{L} = \mathcal{H} - \gamma p \frac{\partial}{\partial p} + \frac{\sigma^2}{2}\frac{\partial^2}{\partial p^2}\,.$$

Letting $V_t = V(q_t,p_t)$ we see that since Ψ solves (15), $\mathcal{A}(P^2)(H_t;1) = \Lambda(H_t)H_t$ and $\mathcal{H}H = 0$, we have that

$$\begin{aligned} dV_t = &-\gamma\Lambda(H_t)H_t dt + \frac{\sigma^2}{2}dt + \sigma p_t dW_t \\ &- \gamma p_t \frac{\partial\Psi}{\partial p}(q_t,p_t;1)dt + \frac{\sigma^2}{2}\frac{\partial^2\Psi}{\partial p^2}(q_t,p_t;1)dt + \sigma\frac{\partial\Psi}{\partial p}(q_t,p_t;1)dW_t\,. \end{aligned} \tag{17}$$

The first line of the right-hand side essentially coincides with (13); therefore, to realize our goal we need to show that this choice of V is comparable to H and that the remaining terms on the right-hand side are negligible at large energies. To accomplish these remaining goals we will exploit the scaling of the equations. To this end we make the following definition:

Definition 1. *We say that a real valued function $\phi(Q,P;\lambda)$ scales like the energy to the power κ if*

$$\phi(\ell^{\frac{2}{\alpha_1}}Q, \ell P;\lambda) = \ell^{2\kappa}\phi(Q,P;\ell\lambda) \tag{18}$$

for all P and Q and $\ell \geq 1$.

Observe that, as expected, H scales like the energy to the power 1 since

$$H(\ell^{\frac{2}{\alpha_1}}Q, \ell P;\lambda) = \ell^2 H\big(Q,P;\ell\lambda\big) \tag{19}$$

Now $(Q,P;\lambda) \mapsto P^2 - \mathcal{A}(P^2)(Q,P;\lambda)$ also scales like the energy to the power one since

$$(\ell P)^2 - \mathcal{A}(P^2)(\ell^{\frac{2}{\alpha_1}}Q,\ell P;\lambda) = \ell^2\Big(P^2 - \mathcal{A}(P^2)(Q,P;\ell\lambda)\Big)\,.$$

We will see that the solution Ψ to (15) scales like the energy to the power $\frac{1}{\alpha_1}+\frac{1}{2}$, which is less than one. Since $|\Psi|$ grows sublinearly in the energy H as $H\to\infty$, $V = H + \Psi$ is comparable to H for large values of H. That is, for any $\epsilon > 0$ there exists a positive C_ϵ so that

$$(1-\epsilon)H - C_\epsilon \leq V \leq (1+\epsilon)H + C_\epsilon\,.$$

The fact that Ψ scales like the energy to the power $\frac{1}{\alpha_1}+\frac{1}{2}$ implies that

$$\begin{aligned} \frac{\partial\Psi}{\partial p}(q,p;1) &= \ell^{\frac{2}{\alpha_1}}\frac{\partial\Psi}{\partial p}\big(\tfrac{q}{\ell^{\frac{2}{\alpha_1}}},\tfrac{p}{\ell};\ell\big) \qquad \text{and} \\ \frac{\partial^2\Psi}{\partial p^2}(q,p;1) &= \ell^{\frac{2}{\alpha_1}-1}\frac{\partial\Psi}{\partial p}\big(\tfrac{q}{\ell^{\frac{2}{\alpha_1}}},\tfrac{p}{\ell};\ell\big)\,. \end{aligned} \tag{20}$$

These observations are made more formal in Proposition 5.5. In light of this scaling, each of the "dt" terms in the second line of (17) can be bounded by $\epsilon H_t + C_\epsilon$ for any $\epsilon > 0$ and some matching constant C_ϵ. From the discussion around the definition of Λ we see that for some positive Λ_0 and C_0, $\Lambda(H_t)H_t \geq 2\Lambda_0 H_t - C_0$. Hence all of the "$dt$" can be bounded by $-\Lambda_0 H_t + C$.

Combining all of these considerations we obtain that for some positive c and C,

$$\frac{d}{dt}\mathbf{E}V_t \leq -c\mathbf{E}V_t + C$$

and thus V is a good Lyapunov function.

This is the basic line of argument we will follow in the rest of text. We will need to establish the scaling and smoothness properties assumed in the preceding calculations. We will also have to modify the argument sightly since while the scaling dictated by (20) controls the terms as the Hamiltonian goes to infinity it presents a problem as the value of the Hamiltonian goes to zero. Namely unless we modify the solution Ψ in a neighborhood of zero the term $\frac{\partial \Psi}{\partial p}$ will explode as $H \to 0$.

5. The Lyapunov function: Rigorous construction

We now repeat the argument outlined in the preceding section while dealing with various technical points neglected previously and addressing the two main deficiencies. First, we will now consider general potentials U of the form given in (1) and not just the special form given in Section 4.1. Second, the careful reader might have noticed that while the scaling given in (20) ensures that the terms in question go to zero at large energy it also implies that they could (depending on the α's) blow up at small energies. Fixing this second point will require a small modification in the argument.

In keeping with the discussion of the previous section we now introduce a family of potentials $U(R;\lambda)$ defined by

$$U(R;\lambda) \stackrel{\text{def}}{=} \frac{a_0}{\lambda^2} + a_1|R|^{\alpha_1} + \sum_{k=2}^{K} a_k |R|^{\alpha_k} \lambda^{2(\frac{\alpha_k}{\alpha_1}-1)} \tag{21}$$

where the a_k and α_k satisfy the same constraints as in (1). Since a_0 can be changed without changing the dynamics, we can assume that $\inf U(R;1) = 0$. Notice that $U(R;1)$ gives the same potential as originally defined in (4). Also notice that $U(\ell^{\frac{2}{\alpha_1}} R;\lambda) = \ell^2 U(R;\ell\lambda)$ for $\ell, R > 0$. Or rather, in the language of Definition 2, U scales as the energy to the power one.

For any $\lambda \geq 1$, consider the one-dimensional Hamiltonian system with Hamiltonian

$$H(Q,P;\lambda) \stackrel{\text{def}}{=} \frac{P^2}{2} + U(Q;\lambda)$$

which scales like the energy to the power one (which if it must if our terminology is at all reasonable). Associated with this Hamiltonian are the dynamics given by

$$\dot{Q}_t = \frac{\partial H}{\partial P}(Q_t,P_t) = P_t \qquad \text{and} \qquad \dot{P}_t = -\frac{\partial H}{\partial Q}(Q_t,P_t) = -U'(Q_t;\lambda)\,.$$

This deterministic system has only closed orbits which lie completely in the upper half plane denoted by $\mathbb{H} = \{(Q,P) \in \mathbb{R}^2 : y > 0\}$ provided $(Q_0,P_0) \in \mathbb{H}$.

We are interested in the above system with the addition of balanced noise and dissipation. Namely, for temperature $T > 0$, friction $\gamma > 0$ and noise variance $\sigma = \sqrt{2\gamma T}$, we have

$$\begin{aligned} dq_t &= p_t\, dt \\ dp_t &= -U'(q_t;\lambda)\, dt - \gamma p_t\, dt + \sigma\, dW_t \end{aligned} \tag{22}$$

The following is the main technical result concerning the existence of a Lyapunov function. The proof follows the argument outlined in Section 4 and will be given in the subsequent sections.

Theorem 5.1. *There exists a C^∞ function $V : \mathbb{H} \to [0,\infty)$ so that for any $\delta > 0$ there exists positive constants c and C with*

$$(1-\delta)H(q,p) - C \le V \le (1+\delta)H(q,p) + C$$

for all $(q,p) \in \mathbb{H}$ and writing $V_t = V(q_t,p_t)$ then one has

$$dV_t \le -\gamma(\Lambda_* - \delta)V_t dt + C dt + dM_t \tag{23}$$

for some L^2-martingale M_t with quadratic variation:

$$\langle M\rangle_t \le (\sigma^2 + \delta)\int_0^t (V_s + C) ds\,.$$

Theorem 5.1 has the following three immediate corollaries.

Corollary 5.2. *Let $\tau_{\mathbb{H}} = \inf\{t \ge 0 : (q_t,p_t) \notin \mathbb{H}\}$; then for all $(p_0,q_0) \in \mathbb{H}$, $\tau_{\mathbb{H}} = \infty$ almost surely. And hence the local in time solutions to (22) for $(p_0,q_0) \in \mathbb{H}$ provided by the standard theory are in fact global in time solutions contained in $\mathbb{H}$ for all time with probability one.*

Proof of Corollary 5.2. For any $R > H(q_0,p_0)$, on the set $D_R = \{(q,p) \in \mathbb{R}^2 : H(q,p) \le R\}$ the coefficients of (22) are Lipschitz. Hence up to the exit time from D_R one has the existence and uniqueness of strong solutions of (22). Defining the stopping time $\tau_R = \inf\{s : (q_s,p_s) \notin D_R\}$, we have constructed strong solutions of (22) for $t < \tau_R$. Since R was arbitrary and the solutions don't depend on R up to time τ_R we need only show that for any $t > 0$,

$$P\Big(\bigcap_{\{k\in\mathbb{Z}:k>H(q_0,p_0)\}} \{\tau_k \le t\}\Big) = 0\,.$$

Now using (23), it is straightforward to see that $\mathbf{E}H_t \le c_0\mathbf{E}V_t < \infty$ for any $t > 0$ and some $c_0 > 0$. This combined with the Doob Inequality for Martingales yields $P(\tau_R \le t) \le P(\sup_{s\le t} H_s \ge R) \le C(t)/R^2$. Since $\sum_{k=1}^\infty k^{-2}$ is summable the result follows from the Borel-Cantelli lemma. □

For any $\beta > 0$, we define $\Phi^{(\beta)}(h) = \exp(\beta h)$ and set $\Phi_t^{(\beta)} = \Phi^{(\beta)}(H_t)$. Then from (23) and the fact that $\sigma^2 = 2\gamma T$, it is clear that the following corollary holds.

Corollary 5.3. *For any $\beta \in (0, 2\Lambda_*/T)$ there exist positive κ and K so that*

$$\mathbf{E}\big[\Phi_{t+s}^{(\beta)}\big|\mathcal{F}_s\big] \le e^{-\kappa t}\Phi_s^{(\beta)} + Kt$$

where $\mathcal{F}_s$ is the σ-algebra generated by $\{(q_r,p_r) : r \le s\}$. Also recall from (12) that $\Lambda_* \in (1,2)$ since $\alpha_1 > 2$ and hence $\beta = \frac{1}{T}$ is always allowed.

For the next Corollary, we momentarily return to considering the unreduced system $(\mathbf{p}_t,\mathbf{q}_t) = (q_1(t),q_2(t),p_1(t),p_2(t))$ defined by (2).

Corollary 5.4. *Let* $\tau_{\mathbb{S}} = \inf\{t \geq 0 : (\mathbf{q}_t, \mathbf{p}_t) \notin \mathbb{S}\}$*; then for all* $(\mathbf{q}_0, \mathbf{p}_0) \in \mathbb{S}$*,* $\tau_{\mathbb{S}} = \infty$ *almost surely. And hence the local in time solutions to* (2) *for* $(\mathbf{q}_0, \mathbf{p}_0) \in \mathbb{S}$ *provided by the standard theory are in fact global in time solutions contained in* $\mathbb{S}$ *for all time with probability one.*

Proof of Corollary 5.4. The existence of a global solution $(\mathbf{q}_t, \mathbf{p}_t)$ to (2) is equivalent the existence of a global solution $(\bar{q}_t, \bar{p}_t, \tilde{q}_t, \tilde{p}_t)$ which solves (3). The existence of a global solution to $(\bar{q}_t, \bar{p}_t)$ follows directly from Corollary 5.2. Since the pair $(\tilde{q}_t, \tilde{p}_t)$ is independent of $(\bar{q}_t, \bar{p}_t)$, we can consider it alone. Since it has no singularity, the existence of a global solution for $(\bar{q}_t, \bar{p}_t)$ can be found in many places including [MSH02]. □

5.1. **The Deterministic Trajectories.** Since the Hamiltonian $H(Q, P; \lambda)$ is conserved for the deterministic dynamics, the deterministic systems will stay on one of the connected components of the level set

$$\mathcal{E}(\eta, \lambda) \stackrel{\text{def}}{=} \left\{(Q, P) : \ \frac{P^2}{2} + U(Q; \lambda) = \eta\right\} \tag{24}$$

for $\eta = H(Q_0, P_0; \lambda)$. For fixed λ, as $\eta \to \infty$ the set $\mathcal{E}(\eta, \lambda)$ approaches the set

$$\left\{(Q, P) : \ \frac{P^2}{2} + a_1|Q|^{\alpha_1} + a_K \lambda^{2(\frac{\alpha_K}{\alpha_1} - 1)}|Q|^{\alpha_K} = \eta\right\} \tag{25}$$

which is of the form discussed in Section 4.1. Observe if we define the scaling map $S_\ell \colon (Q, P) \mapsto (\ell^{\frac{2}{\alpha_1}} Q, \ell P)$ then from the scaling of the terms in H, one has

$$\mathcal{E}(\eta, \lambda) = S_\ell\big(\mathcal{E}(\ell^{-2}\eta, \ell\lambda)\big) .$$

Hence the as $\lambda \to \infty$, one also has that the set $\mathcal{E}(\eta, \lambda)$ converges to the set in (25).

Hence for either fixed λ and large η or fixed η and large λ, $\mathcal{E}(\eta, \lambda)$ is homotopic to a circle and can be written as the union of two segments which are graphs over the P axis (see Figure 1). For η large enough, increasing λ does not destroy this property. That is to say there exists an η_* so that if $\eta \geq \eta_*$ and $\varrho(Q; \eta, \lambda) = \sqrt{2(\eta - U(Q; \lambda))}$ then

$$\left\{\big(Q, \varrho(Q; \eta, \lambda)\big), \big(Q, -\varrho(Q; \eta, \lambda)\big) : Q \in [Q_-(\eta, \lambda), Q_+(\eta, \lambda)]\right\} = \mathcal{E}(\eta, \lambda)$$

and $\mathcal{E}(\eta, \lambda)$ consists of a single connected component which is topologically equivalent to a circle for all $\lambda \geq 1$. Here

$$\begin{aligned} &Q_-(\eta, \lambda) \text{ is the smallest solution of } \eta - U(Q; \lambda) = 0 \\ &Q_+(\eta, \lambda) \text{ is the largest solution of } \eta - U(Q; \lambda) = 0 \end{aligned} \tag{26}$$

Lastly as discussed in Section 4.1, for fixed η, as $\lambda \to \infty$ we have that $\mathcal{E}\eta, \lambda)$ converges pointwise to

$$\left\{(Q, P) : \ \frac{P^2}{2} + a_1|Q|^{\alpha_1} = \eta \text{ and } Q > 0\right\} .$$

5.2. **Averaging along trajectories.** We now consider the value of certain functions averaged along trajectories of the deterministic system. Because we will only need the averages of functions at large η or λ, we will restrict to functions $\phi \in \mathbb{H}_0$ where

$$\mathbb{H}_0 \stackrel{\text{def}}{=} \left\{\phi : \mathbb{H} \times [1, \infty) \to \mathbb{R} : \phi(Q, P; \lambda) = 0 \text{ if } H(Q, P; \lambda) < \eta_*\right\}$$

where η_* is the constant from Section 5.1.

Since we will only be averaging functions $\phi \in \mathbb{H}_0$, we need only to average along $\mathcal{E}(\eta, \lambda)$ which are connected and topologically a circle as discussed in the last section.

On such $\mathcal{E}$ observe that if $\tau(\eta, \lambda)$ is the period of the orbit then $P_t = \varrho(Q_t)$ for $t \in [0, \tau/2]$ and $P_t = -\varrho(Q_t)$ for $t \in [\tau/2, \tau]$ if $(Q_0, P_0) = (Q_-, 0)$. And we define

$$\begin{aligned}\langle\phi\rangle(\eta,\lambda) &= \int_0^{\tau(\eta,\lambda)} \phi(Q_t, P_t; \lambda)dt \\ &= \int_0^{\frac{\tau(\eta,\lambda)}{2}} \phi(Q_t, \varrho(Q_t); \lambda)dt + \int_{\frac{\tau(\eta,\lambda)}{2}}^{\tau(\eta,\lambda)} \phi(Q_t, -\varrho(Q_t); \lambda)dt \\ &= \int_{Q_-(\eta,\lambda)}^{Q_+(\eta,\lambda)} \frac{\phi(q, \varrho(q;\eta,\lambda);\lambda) + \phi(q, -\varrho(q;\eta,\lambda);\lambda)}{\varrho(q;\eta,\lambda)} dq\end{aligned}$$

The period τ can be obtained by $\tau(\eta, \lambda) = \langle 1\rangle(\eta, \lambda)$. We define the "averaging operator" $\mathcal{A}$ by

$$\mathcal{A}(\phi)(\eta,\lambda) = \frac{1}{\tau(\eta,\lambda)}\langle\phi\rangle(\eta,\lambda)\,.$$

Note that if $\Psi(Q, P; \lambda) = \psi(H(Q, P; \lambda); \lambda)$ then $\mathcal{A}(\Psi)(\eta, \lambda) = \psi(\eta, \lambda)$ since the energy is constant on an orbit.

5.3. **Scaling across trajectories.** We will need the following refined notion of the scaling of a function which allows for the scaling to hold only for sufficiently large energy.

Definition 2. *We say that the real valued function $\phi(Q, P; \lambda)$* EVENTUALLY SCALES *like the energy to the κ if there exists a constant η_ϕ such that* (18) *holds for all (Q, P) with $H(Q, P; \lambda) \geq \eta_\phi$ and $\lambda \geq 1$.*

Proposition 5.5. *If $\phi \in \mathbb{H}_0$ (eventually) scales like the energy to the power κ then:*

$$\begin{aligned}\langle\phi\rangle(Q,P;\lambda) &\text{ (eventually) scales like the energy to the power } \kappa + \frac{1}{\alpha_1} - \frac{1}{2}\,,\\ \mathcal{A}(\phi)(Q,P;\lambda) &\text{ (eventually) scales like the energy to the power } \kappa\,,\\ \frac{\partial\phi}{\partial P}(Q,P;\lambda) &\text{ (eventually) scales like the energy to the power } \kappa - \frac{1}{2}\,,\\ \frac{\partial\phi}{\partial Q}(Q,P;\lambda) &\text{ (eventually) scales like the energy to the power } \kappa - \frac{1}{\alpha_1}\,.\end{aligned}$$

In particular, the period of the determinisitc orbits $\tau(\eta, \lambda)$ stasifies $\tau(h\eta, \lambda) = h^{\frac{1}{\alpha_1}-\frac{1}{2}}\tau(\eta, h^{\frac{1}{2}}\lambda)$ for $\eta \geq \eta_*$ and $\lambda \geq 1$, and hence; the function $(Q, P; \lambda) \mapsto \tau(H(Q, P; \lambda), \lambda)$ eventually scales like the energy to the power $\frac{1}{\alpha_1} - \frac{1}{2}$.

Proof. Recalling the definition of $Q_\pm$ from (26) and ϱ from (9), notice that

$$\begin{aligned}Q_\pm(h\eta,\lambda) &= h^{\frac{1}{\alpha_1}}Q_\pm(\eta, h^{\frac{1}{2}}\lambda)\\ h^{\frac{1}{2}}\varrho(q;\eta,h^{\frac{1}{2}}\lambda) &= \varrho(h^{\frac{1}{\alpha_1}}q; h\eta, \lambda)\end{aligned}$$

We begin with $\langle\phi\rangle$ and $\mathcal{A}(\phi)$. By the assumption on ϕ, $\phi(h^{\frac{1}{2}}p, h^{\frac{1}{\alpha_1}}q;\lambda) = h^{\kappa}\phi(q,p;h^{\frac{1}{2}}\lambda)$, so

$$\begin{aligned}\langle\phi\rangle(h\eta,\lambda) &= \int_{Q_-(h\eta,\lambda)}^{Q_+(h\eta,\lambda)} \frac{\phi(q,\varrho(q;h\eta,\lambda));\lambda) + \phi(q,-\varrho(q;h\eta,\lambda))\lambda)}{\varrho(q;h\eta,\lambda)}dq \\ &= \int_{Q_-(\eta,h^{\frac{1}{2}}\lambda)}^{Q_+(\eta,h^{\frac{1}{2}}\lambda)} \frac{\phi(h^{\frac{1}{\alpha_1}}q, h^{\frac{1}{2}}\varrho(q;\eta,h^{\frac{1}{2}}\lambda));\lambda) + \phi(h^{\frac{1}{\alpha_1}}q, -h^{\frac{1}{2}}\varrho(q;\eta,h^{\frac{1}{2}}\lambda));\lambda)}{h^{\frac{1}{2}-\frac{1}{\alpha_1}}\varrho(q;\eta,h^{\frac{1}{2}}\lambda)}dq \\ &= h^{\kappa+\frac{1}{\alpha_1}-\frac{1}{2}}\langle\phi\rangle(\eta,h^{\frac{1}{2}}\lambda)\end{aligned}$$

Since $(q,p)\mapsto 1$ scales like the energy to the 0 power, we see that

$$\tau(h\eta,\lambda) = \langle 1\rangle(h\eta,\lambda) = h^{\frac{1}{\alpha_1}-\frac{1}{2}}\langle 1\rangle(\eta,h^{\frac{1}{2}}\lambda) = h^{\frac{1}{\alpha_1}-\frac{1}{2}}\tau(\eta,h^{\frac{1}{2}}\lambda)$$

and hence $\mathcal{A}(\phi)(h\eta,\lambda) = h^{\kappa}\mathcal{A}(\phi)(\eta,h^{\frac{1}{2}}\lambda)$.

We now turn to the statements about the derivatives. If we assume that ϕ (eventually) scales like the energy to the power κ, differentiating both sides of (18)

$$\begin{aligned}\ell\frac{\partial\phi}{\partial P}(\ell^{\frac{2}{\alpha_1}}Q,\ell P;\lambda) &= \ell^{2\kappa}\frac{\partial\phi}{\partial P}(Q,P;\ell\lambda) \\ \ell^{\frac{2}{\alpha_1}}\frac{\partial\phi}{\partial Q}(\ell^{\frac{2}{\alpha_1}}Q,\ell P;\lambda) &= \ell^{2\kappa}\frac{\partial\phi}{\partial Q}(Q,P;\ell\lambda)\end{aligned}$$

and hence $\frac{\partial\phi}{\partial P}$ and $\frac{\partial\phi}{\partial Q}$ (eventually) scale as the energy to the power $\kappa-\frac{1}{2}$ and $\kappa-\frac{1}{\alpha_1}$ respectively. □

5.4. **New Definition of Λ and its Control.** In the heuristic discussion of Section 4.1, we defined Λ in terms of the average of P^2 at energy one with an appropriately scaled parameter λ. We were free to use energy one as our reference energy since the simple potential used in Section 4.1 has simply connected level-sets at all energies. For the more complicated potential defined in (21), this is only guaranteed for energy η greater than the η_* introduced in Section 5.1. Hence we define

$$\Lambda(\eta) = \begin{cases}\frac{1}{\eta_*}\mathcal{A}(P^2)(\eta_*;\left(\frac{\eta}{\eta_*}\right)^{\frac{1}{2}}) & \text{for } \eta\geq\eta_* \\ 0 & \text{for } \eta<\eta_*\end{cases} \tag{27}$$

With this definition we have that

$$\mathcal{A}(P^2)(\eta,1) \geq \eta\Lambda(\eta) \text{ for all } \eta\geq 0 \tag{28}$$

and that $\Lambda(\eta)\to\Lambda_*$ as $\eta\to\infty$. Here Λ_* is as defined in (12) and does not depend on our choice of η_*. In fact the following proposition which we will need later follows directly from these facts

Lemma 5.6. *For any $\delta>0$ there exists a $C_\delta>0$ so that*

$$(\Lambda_*-\delta)\eta - C_\delta \leq \eta\Lambda(\eta) \leq (\Lambda_*+\delta)\eta + C_\delta \tag{29}$$

for all $\eta\geq 0$.

5.5. **The associated Poisson equation.** Recalling the definition of the Liouville operator $\mathcal{H}$ from (14), given a $\Phi \in \mathbb{H}_0$ we will wish to solve

$$\mathcal{H}\Psi = \Phi - \mathcal{A}(\Phi)\,. \tag{30}$$

Notice that by subtracting $\mathcal{A}(\Phi)$ we ensure that the right-hand side is in the kernel of the adjoint of $\mathcal{H}$. This is necessary to ensure solvability of equation (30).

Proposition 5.7. *Let $\Phi \in \mathbb{H}_0 \cap C^k(\mathbb{H})$ for some $k \geq 0$. Then* (30) *has a solution in $\mathbb{H}_0 \cap C^k(\mathbb{H})$.*

Proof. We recall that since we have restricted to ϕ which are zero for energies less than η_*; hence, we need only solve the equation in regions where the level sets are simply connected and homotopically equivalent to a circle.

In other words, all of the characteristic curves of the operator $\mathcal{H}$ are closed curves. Hence one can construct a solution by simply specifying values of Ψ transverse to the characteristic curves, say $\Psi = 0$ on (q, p) with $p = 0$ and $H(q, p) \geq \eta_*$. The remaining values of Ψ are obtained by integrating $\Phi - \mathcal{A}(\Phi)$ around the orbit in, say, the clockwise direction. This is well defined since by construction $\Phi - \mathcal{A}(\Phi)$ has total integral 0 around an orbit and hence the values of Ψ from above and below $p = 0$ will match. Since for $\eta \geq \eta_*$, we know the orbits are simple loops the integration is straightforward along the lines of the previous section on averaging.

The smoothness and local boundedness are direct consequences of the same properties of Φ and the fact that we have excluded the possible singularity at (q, p) with $H(q, p) = 0$ by requiring that Φ be zero in a neighborhood of the point. □

Proposition 5.8. *If Φ (eventually) scales like the energy to the power κ and Ψ solves* (30) *then Ψ (eventually) scales like the energy to the power $\kappa - \frac{1}{2} + \frac{1}{\alpha_1}$.*

Proof. Since H scales like the energy to the first power, Proposition 5.5 implies that $\frac{\partial \Psi}{\partial P}$ and $\frac{\partial \Psi}{\partial Q}$ scale like the energy to the power $\frac{1}{2}$ and $1 - \frac{1}{\alpha_1}$ respectively. Similarly if we assume that Ψ (eventually) scales like the energy to the power ν, then $\mathcal{L}\Psi$ scales like the energy to the power $\nu + \frac{1}{2} - \frac{1}{\alpha_1}$. If Φ scales like the energy to the power κ then Proposition 5.5 implies that $\Psi = \Phi - \mathcal{A}(\Phi)$ also scales to the power κ. Hence if $\mathcal{L}\Psi = \Phi - \mathcal{A}(\Phi)$, we conclude that $\kappa = \nu + \frac{1}{2} - \frac{1}{\alpha_1}$ and therefore $\nu = \kappa - \frac{1}{2} + \frac{1}{\alpha_1}$ as claimed. □

Remark 5.9. *Since we have assumed $\alpha_1 > 2$, the solution to the Poisson equation* (30) *scales with a lower power than the right-hand side of* (30) *. See the discussion in Section 8 for further perspective on the case $\alpha_1 = 2$.*

5.6. **Building the Lyapunov function: Proof of Theorem 5.1.** Let χ be a monotone increasing, C^∞ function with all of its derivatives bounded such that $\chi(\eta) = 0$ for $\eta < \eta_*$ and $\chi(\eta) = 1$ for $\eta \geq \eta_* + 1$ where η_* was used in the definition of $\mathbb{H}_0$. We define $\chi(Q, P; \lambda) = \chi(H(Q, P; 1))$. Then $\Upsilon(Q, P; \lambda) = P^2\chi(Q, P; \lambda) \in \mathbb{H}_0$ since $\Upsilon(Q, \ell P; \lambda) = \ell^2 P^2 \chi(Q, P; \lambda)$ for $H \geq \eta_* + 1$.

By construction $\Upsilon \in \mathbb{H}_0$ and is also smooth and locally bounded; and hence, Proposition 5.7 ensures the existence of a smooth locally bounded solution to

$$(\mathcal{H}\Psi)(Q, P; \lambda) = \gamma\Big(\Upsilon - \mathcal{A}(\Upsilon)(Q, P; \lambda)\Big)\,. \tag{31}$$

As before we define $V_t = H_t + \Psi_t$ and observe that

$$\begin{aligned} dV_t &= -\gamma \mathcal{A}(\Upsilon)(H_t;1)dt + \frac{\sigma^2}{2}dt + \sigma(p_t + \frac{\partial \Psi}{\partial P}(q_t,p_t))dW_t \\ &\qquad - \gamma p_t \frac{\partial \Psi}{\partial P}(q_t,p_t)dt + \frac{\sigma^2}{2}\frac{\partial^2 \Psi}{\partial P^2}(q_t,p_t)dt + \gamma(\Upsilon(q_t,p_t) - p_t^2)dt \\ &\leq -\gamma \Lambda(H_t)H_t dt + \frac{\sigma^2}{2}dt + E_t dt + G_t dt + dM_t \end{aligned}$$

where $E_t = E(q_t,p_t)$, $G_t = G(q_t,p_t)$ and

$$E(q,p) = -\gamma p \frac{\partial \Psi}{\partial P}(q,p) + \frac{\sigma^2}{2}\frac{\partial^2 \Psi}{\partial P^2}(q,p) \quad \text{and} \quad G(q,p) = \gamma(\Upsilon(q,p) - p^2).$$

The martingale M_t is defined by

$$M_t = \sigma \int_0^t \Big(p_s + \frac{\partial \Psi}{\partial P}(q_s,p_s)\Big)dW_s\,.$$

To prove the first claim, it is enough to show that for any $\delta_0, \delta_1 > 0$, $\delta_0\Psi_t + E_t + G_t \leq \delta_1 H_t + C_\epsilon$ for some constant C_ϵ. First observe that $G_t \leq 0$ for all $t \geq 0$, so we need only show that $\delta_0\Psi_t + E_t \leq \epsilon H_t + C_\epsilon$. Since $\Upsilon(Q,P;\lambda)$ eventually scales like the energy, Proposition 5.8 implies that Ψ eventually scales like the energy to the power $\frac{1}{2} + \frac{1}{\alpha_1}$ which by assumption is less than one. Proposition 5.5 in turn implies that $\partial_P \Psi$ eventually scales like the energy to the power $\frac{1}{\alpha_1}$ and $\partial_P^2 \Psi$ eventually scales like the energy to the power $\frac{1}{\alpha_1} - \frac{1}{2}$ which is again less than one. Hence, we conclude that both Ψ and E are dominated by a function which scales like the energy with a power strictly less than one. (Notice that both E_t and Ψ_t are identically zero if $H_t \leq \eta_*$.) Using this scaling, we conclude that $\delta_0\Psi_t + E_t$ has the desired bound. Combining these observations with (29), we see that for any $\delta > 0$ there exist a positive C so that

$$dV_t \leq -\gamma(\Lambda_* - \delta)V_t dt + C dt + dM_t\,.$$

From the scaling it is clear that for any $\delta > 0$ there exists a positive C such that

$$(1-\delta)H(q,p) - C \leq V(q,p) \leq (1+\delta)H(q,p) + C$$

for all $(q,p) \in \mathbb{H}$, which completes the proof of Theorem 5.1. Letting $\langle M \rangle_t$ denote the quadratic variation of the continuous martingale M_t, following arguments similar to those just complete, we see that $d\langle M \rangle_t \leq (\frac{\sigma^2}{2} + \delta)(V_t + C)dt$.

6. Existence of a "small set"

The following Lemma provides the remaining missing ingredient required to prove geometric ergodicity.

Lemma 6.1. *For every $\eta > 0$, there exists a probability measure ν supported in $\mathbb{H}$, a $t > 0$ and $c_0 > 0$ so that*

$$\inf_{(q,p):H(q,p)\leq\eta} \mathcal{P}_t((q,p),A) \geq c_0\nu(A)$$

for any $A \subset \mathbb{H}$.

Proof. Let $\mathcal{L}$ denote the generator of the Markov semigroup associated to (22). We begin by observing that for any $s \in (0,t)$ the operator $\partial_s - \mathcal{L}$ is hypoelliptic. (See [MSH02] for the straightforward calculation of Lie-brackets.) Fixing an $x_0 = (p_0, q_0)$, this implies that the transition $\mathcal{P}_s(x_0, dy)$ to $y = (q_s, p_s)$ has a smooth density inside of $\mathbb{H}$. In particular, there exists a $\rho_s(x_0, y)$ so that

$$\mathcal{P}_s(x_0, B_\delta(x_0)) = \int_{B_\delta(x_0)} \rho_s(x_0, y) dy$$

for a sufficiently small δ-ball around x_0 and a sufficiently small $s > 0$. Furthermore, $\rho_s(x_0, y)$ is smooth at $y \in B_\delta(x_0)$. By considering the adjoint of $\mathcal{L}$, one can conclude that $\rho_s(x_0, y)$ is smooth at $(x_0, y) \in B_\delta(x_0) \times B_\delta(x_0)$. (See [Str08]). Since for small enough t we are sure that $\mathcal{P}_s(x_0, B_\delta(x_0)) > 0$ there must exist a $y_0 \in B_\delta(x_0)$, a $c_0' > 0$ and a possibly smaller δ so that

$$\inf_{(x,y) \in B_\delta(x_0) \times B_\delta(y_0)} \rho_s(x, y) \geq c_0' > 0 \,.$$

Now one can follow Lemma 3.4 of [MSH02] to construct a control argument ensuring that given any open set $\mathcal{O} \subset \mathbb{H}$ and $\mathbb{H}(\eta) = \{(q,p) : H(q,p) \leq \eta\}$ there exists a $t > 0$ and c_0'' such that

$$\inf_{z \in \mathbb{H}(\eta)} \mathcal{P}_t(z, \mathcal{O}) \geq c_0''$$

The argument in [MSH02] assumes that the drift vector field is bounded on compact sets. This is still true if we restrict to $\mathbb{H}(\eta)$ for any finite $\eta > 0$. The uniform lower bound is not explicitly mentioned, however one can pick a single tubular neighborhood size of the needed control and ensure that the control and its derivatives are uniformly bounded for all starting and ending points in $\mathbb{H}(\eta)$.

Setting $t = r + s$, defining ν as normalized Lebesgue measure on $B_\delta(y_0)$ and combining the preceding two estimates produces, for any $z \in \mathbb{H}(\eta)$ and $A \subset \mathbb{H}$

$$\begin{aligned}
\mathcal{P}_t(z, A) &= \int_{\mathbb{H}} \mathcal{P}_r(z, dy) \mathcal{P}_s(y, A) \\
&\geq \int_{B_\delta(x_0)} \mathcal{P}_r(z, dy) \mathcal{P}_s(y, A \cap B_\delta(y_0)) \\
&\geq \int_{A \cap B_\delta(y_0)} \int_{B_\delta(x_0)} \mathcal{P}_r(z, dy) \mathcal{P}_s(y, dz) \\
&\geq c_0'' \lambda_{leb}(A \cap B_\delta(y_0)) \int_{B_{\delta(x_0)}} \mathcal{P}_r(z, dy) \\
&\geq c_0' c_0'' \lambda_{leb}(A \cap B_\delta(y_0)) = \frac{c_0' c_0''}{\lambda_{leb}(B_\delta(y_0))} \nu(A)
\end{aligned}$$

which concludes the proof. □

7. Proof of Theorem 3.1

Theorem 3.1 follows by combining Theorem 5.1 and Lemma 6.1 and invoking Theorem 1.2 from [HM08]. This result is a repackaging of a well known result of Harris. It can be found in many places. Most appropriate for the current discussion is the work of Meyn and Tweedie exemplified by [MT93, Section 15].

8. Conclusion

We began by observing that to leading order, the dynamics at high energy follow the deterministic dynamics given by a modified Hamiltonian perturbed by a small noise. To leverage this observation stochastic averaging techniques, built on auxiliary Poisson equation methods, were used to construct a Lyapunov function sufficient to prove exponential convergence to equilibrium. The central result, given in Theorem 3.1, covers important singular potentials, including Lennard-Jones type potentials, which had not been covered by previous results. Theorem 3.1 has two principal remaining deficiencies. First it only applies to two interacting particles in isolation. Second, Theorem 3.1 does not cover the classical case where the confining potential U grows quadratically at infinity.

In principle, the extension to many particles could follow a similar route, since when two particles are near each other their principal interaction is with each other while other particles are just a small perturbation. However it is possible that the orbit over which one must average could also interact with other particles. This would make finding closed form representations of the averaging measure difficult at best (chaotic orbits are to be expected). Even if in some setting the high energy orbits remain of the type considered here, the combinatorics of the possible interactions would be complicated.

In contrast, the extension to potentials with quadratic growth is almost certainly within reach. In fact, Figure 3 gives a strong indication how to proceed. Since for $\alpha_1 = 2$ the period of oscillation is not going to zero as the energy of the system increases, instantaneous homogenization/averaging of the effect of one orbit is not feasible. However, building on an idea from [RBT02] one could consider the average of the energy over one period τ of the system. First observe that τ has a limit $\tau_* > 0$ as the energy goes to ∞. Namely one would consider the quantity

$$V_t = \frac{1}{\tau_*} \int_{t-\tau_*}^{t} H_s ds \,.$$

Then using (7) one obtains

$$\frac{\partial V_t}{\partial t} = \frac{1}{\tau_*}\big(H_t - H_{t-\tau_*}\big) = -\frac{\gamma}{\tau_*} \int_{t-\tau_*}^{t} p_s^2 ds + \frac{\sigma^2}{2} + \frac{\sigma}{\tau_*} \int_{t-\tau_*}^{\tau_*} p_s dW_s \,.$$

Since at high energy p_s will be very close to the deterministic orbit, one can likely prove that

$$\frac{1}{\tau_*}\mathbf{E} \int_{t-\tau_*}^{t} p_s^2 ds \approx \frac{1}{\tau_*} \int_{t-\tau_*}^{t} \mathbf{E}\big[\Lambda(H_s) H_s\big] ds \approx \Lambda_* \mathbf{E} V_t \,.$$

This could then be used to obtain control of the excursions away from the center of space. Note that following the above argument will not produce an Lyapunov function which infinitesimally decreasing on average as was constructed in this note. This argument essentially amortizes the total energy dissipation which occurs over a single orbit, smoothing out the times when the infinitesimal rate of energy dissipation nears zero. Since we feel that covering the quadratic case is not sufficient motivation for the extra complications, we have chosen to use the approach used in this note since it leads to a more classical "infinitesimal" Lyapunov function.

Acknowledgments. The authors wishes to thank the NSF for its support through grants DMS-0449910 (JCM), DMS-0854879 (JCM) and DMS-0204690 (SCS).

JCM would like to thank Martin Hairer and Luc Rey-Bellet for interesting and informative discussions.

References

[AT87] M. P. Allen and D. J. Tildesley. Computer Simulation of Liquids. Oxford, 1987.

[BBO+83] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comp. Chem., 4:187–217, 1983.

[CS08] Ben Cooke and Scott C. Schmidler. Preserving the Boltzmann ensemble in replica-exchange molecular dynamics. J. Chem. Phys., 129(16):164112–17), 2008.

[FS96] Daan Frenkel and Berend Smit. Understanding Molecular Simulation. Academic Press, 1996.

[HM08] Martin Hairer and Jonathan C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. arXiv/0810.2777, 2008.

[HM09] Martin Hairer and Jonathan C. Mattingly. Slow energy dissipation in anharmonic oscillator chains. Comm. Pure Appl. Math., 62(8):999–1032, 2009.

[IL71] I. A. Ibragimov and Yu. V. Linnik. Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, 1971.

[JTR88] W. L. Jorgensen and J. Tirado-Rives. The OPLS force field for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J. Amer. Chem. Soc., 110:1657–1666, 1988.

[Ken04] W. S. Kendall. Geometric ergodicity and perfect simulation. Elec.Comm. Prob., 9:140–151, 2004.

[Lea96] Andrew R. Leach. Molecular Modelling: Principles and Applications. Addison Wesley Longman Ltd., 1996.

[MSH02] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. Stochastic Process. Appl., 101(2):185–232, 2002.

[MT93] S. P. Meyn and R. L. Tweedie. Markov chains and stochastic stability. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.

[Pas94] R. W. Pastor. Techniques and applications of Langevin dynamics simulations. In G. R. Luckhurst and C. A. Veracini, editors, The Molecular Dynamics of Liquid Crystals, pages 85–138. Kluwer Academic, 1994.

[PCC+95] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. R. Ross, III T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman. AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. Comp. Phys. Commun., 91:1–41, 1995.

[RBT02] Luc Rey-Bellet and Lawrence E. Thomas. Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. Comm. Math. Phys., 225(2):305–329, 2002.

[Sch02] Tamar Schlick. Molecular Modeling and Simulation. Springer-Verlag, 2002.

[Str08] Daniel W. Stroock. Partial differential equations for probabilists, volume 112 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2008.

Department of Mathematics, Duke University, Durham NC

Department of Mathematics, Center for Theoretical and Mathematical Sciences, Center for Nonlinear and Complex Systems, and Department of Statistical Sciences, Duke University, Durham, NC, 27708-0251

Department of Mathematics, Duke University, Durham NC and Department of Mathematics, University of Florida, Gainseville, FL

Department of Statistical Science, Department of Computer Science, Program in Computational Biology and Bioinformatics, Program in Structural Biology and Biophysics, Duke University, Durham NC 27708